

IN THE CLAIMS

1. A process for determining server performance metrics in a network, comprising:
 - providing a service metric probe resident on a server for determining service availability and metric measurements of types of services provided by a content delivery server;
 - providing a latency probe resident on a server for determining latency values for various content delivery servers within the network;
 - wherein the service metric probe consults a configuration file containing each DNS name in an area associated with the service metric probe and any set(s) of services associated with each DNS name;
 - wherein the set(s) of services include any of: HTTP, HTTPS, FTP, streaming media, or generic SNMP;
 - wherein the latency probe calculates a latency value from a location of the latency probe to a client's location; and
 - wherein, for a given DNS name, a DNS server uses updates to the latency values and updates to the service availability and metric measurements to determine a content delivery server to return.
2. The process of claim 1, wherein each content delivery server in the network has a metric test associated with each service supported by each content delivery server.

3. The process of claim 1, wherein the service metric probe periodically performs metric tests on content delivery servers within the service metric probe area, and wherein the service metric probe means records metric results from the periodic tests.
4. The process of claim 1, wherein the latency probe calculates a round trip time for sending a packet to a client to obtain the latency value, and wherein round trip time tests that the latency probe performs, includes any of: PING, UDP Reverse Name lookup, or UDP Packets to high number ports.
5. The process of claim 1, wherein when the latency probe sends a UDP Packet probe to high number ports that fail, the latency probe resends the UDP Packet probe with a low TTL number and increments the TTL number until failure occurs, a last successful TTL number indicates partial latency data.
7. The process of claim 1, wherein the service metric probe means sends an update to all DNS servers in the network that consists of all tests since a last update.
8. (Currently Amended) The process of claim 1, wherein the latency probe updates DNS servers with [[a]] client[[s']] latency data.
9. The process of claim 1, wherein a DNS server uses latency data updates from the latency probe to determine a closest content delivery server to a client.

11. The process of claim 1, wherein the service metric probe sends a packet request to a content delivery server and receives, in response, a packet containing various metrics of the content delivery server, and wherein the service metric probe combines the content delivery server's metrics to arrive at a load metric which is sent to at least one DNS server.

12. (Currently Amended) A process for a determining server performance metrics in a network, comprising the steps of:

providing service metric probe resident on a server for determining service availability and metric measurements of types of services provided by a content delivery server;

providing latency probe resident on a server for determining a latency value for various servers within the network;

wherein the service metric probe sends an update to all DNS servers in the network that consists of all service availability and metric measurements since a last update;

wherein the latency probe updates the DNS servers with client[[s']] latency data; and

wherein, for a given DNS name, a DNS server uses updates to the latency values and updates to the service availability and metric measurements to determine a content delivery server to return.

13. The process of claim 12, wherein the service metric probe consults a configuration file containing each DNS name in the service metric probe area and any set(s) of services associated with each DNS name, and wherein the services include any of: HTTP, HTTPS, FTP, streaming media, or generic SNMP.

14. The process of claim 12, wherein the latency probe calculates a latency value from the latency probe location to a client's location.

15. The process of claim 12, wherein each content delivery server in the network has a metric test associated with each service supported by the content delivery server.

16. The process of claim 12, wherein the service metric probe periodically performs metric tests on content delivery servers within the service metric probe area, and wherein the service metric probe means records metric results from the periodic tests.

17. The process of claim 12, wherein the latency probe calculates a round trip time for sending a packet to a client to obtain latency value, and wherein round trip time tests that the latency probe performs, includes any of: PING, UDP Reverse Name lookup, or UDP Packets to high number ports.

18. The process of claim 12, wherein when the latency probe sends a UDP Packet probe to high number ports that fail, the latency probe resends the UDP Packet probe with a low TTL number and increments the TTL number until failure occurs, a last successful TTL number indicates partial latency data.

19. The process of claim 12, wherein a DNS server uses the latency data updates to determine a closest content delivery server to a client.

21. The process of claim 12, wherein the service metric probe sends a packet request to a content delivery server and receives, in response, a packet containing the various metrics of the content delivery server, and wherein the service metric probe combines the content delivery server metrics to arrive at a load metric which is sent to the DNS servers.

22. A program storage medium readable by a computer, tangibly embodying a program of instructions executable by the computer to perform method steps for determining server performance metrics in a network, comprising the steps of:

providing a service metric probe resident on a server for determining service availability and metric measurements of types of services provided by a content delivery server;

providing a latency probe resident on a server for determining latency values for various content delivery servers within the network;

wherein the service metric probe consults a configuration file containing each DNS name in an area associated with the service metric probe and any set(s) of services associated with each DNS name;

wherein the set(s) of services include any of: HTTP, HTTPS, FTP, streaming media, or generic SNMP;

wherein the latency probe calculates a latency value from a location of the latency probe to a client's location; and

wherein, for a given DNS name, a DNS server uses updates to the latency values and updates to the service availability and metric measurements to determine a content delivery server to return.

23. The method of claim 22, wherein each content delivery server in the network has a metric test associated with each service supported by each content delivery server.

24. The method of claim 22, wherein the service metric probe periodically performs metric tests on content delivery servers within the service metric probe area, and wherein the service metric probe records metric results from the periodic tests.

25. The method of claim 22, wherein the latency probe calculates a round trip time for sending a packet to a client to obtain the latency value, and wherein round trip time tests that the latency probe performs, includes any of: PING, UDP Reverse Name lookup, or UDP Packets to high number ports.

26. The method of claim 22, wherein when the latency probe sends a UDP Packet probe to high number ports that fail, the latency probe resends the UDP Packet probe with a low TTL number and increments the TTL number until failure occurs, a last successful TTL number indicates partial latency data.

28. The method of claim 22, wherein the service metric probe sends an update to all DNS servers in the network that consists of all tests since a last update.

29. (Currently Amended) The method of claim 22, wherein the latency probe updates DNS servers with [[a]] client[[s']] latency data.

30. The method of claim 22, wherein a DNS server uses latency data updates to determine a closest content delivery server to a client.

32. The method of claim 22, wherein the service metric probe sends a packet request to a content delivery server and receives, in response, a packet containing various metrics of the content delivery server, and wherein the service metric probe combines the content delivery server's metrics to arrive at a load metric which is sent to at least one DNS server.

33. (Currently Amended) A program storage medium readable by a computer, tangibly embodying a program of instructions executable by the computer to perform method steps for a determining server performance metrics in a network, comprising the steps of:

providing service metric probe resident on a server for determining service availability and metric measurements of types of services provided by a content delivery server;

providing latency probe resident on a server for determining a latency value for various servers within the network;

wherein the service metric probe sends an update to all DNS servers in the network that consists of all service availability and metric measurements since a last update;

wherein the latency probe updates the DNS servers with client[[s']] latency data; and

wherein, for a given DNS name, a DNS server uses updates to the latency values and updates to the service availability and metric measurements to determine a content delivery server to return.

34. The method of claim 33, wherein the service metric probe consults a configuration file containing each DNS name in the service metric probe area and any set(s) of services associated with each DNS name, and wherein the services include any of: HTTP, HTTPS, FTP, streaming media, or generic SNMP.

35. The method of claim 33, wherein the latency probe calculates a latency value from the latency probe location to a client's location.

36. The method of claim 33, wherein each content delivery server in the network has a metric test associated with each service supported by the content delivery server.

37. The method of claim 33, wherein the service metric probe periodically performs metric tests on content delivery servers within the service metric probe area, and wherein the service metric probe means records metric results from the periodic tests.

38. The method of claim 33, wherein the latency probe calculates a round trip time for sending a packet to a client to obtain a latency value, and wherein round trip time tests that the latency probe performs, includes any of: PING, UDP Reverse Name lookup, or UDP Packets to high number ports.

39. The method of claim 33, wherein when the latency probe sends a UDP Packet probe to high number ports that fail, the latency probe resends the UDP Packet probe with a low TTL number and increments the TTL number until failure occurs, a last successful TTL number indicates partial latency data.

40. The method of claim 33, wherein a DNS server uses the latency data updates to determine a closest content delivery server to a client.

42. The method of claim 33, wherein the service metric probe sends a packet request to a content delivery server and receives, in response, a packet containing the various metrics of the content delivery server, and wherein the service metric probe combines the content delivery server metrics to arrive at a load metric which is sent to the DNS servers.

43. An apparatus for a determining server performance metrics in a network, comprising:
 a service metric probe resident on a server for determining service availability and metric measurements of types of services provided by a content delivery server;

a latency probe resident on a server for determining latency values for various content delivery servers within the network;

 wherein the service metric probe consults a configuration file containing each DNS name in an area associated with the service metric probe and any set(s) of services associated with each DNS name;

wherein the set(s) of services include any of: HTTP, HTTPS, FTP, streaming media, or generic SNMP;

wherein the latency probe calculates a latency value from a location of the latency probe to a client's location; and

wherein, for a given DNS name, a DNS server uses updates to the latency values and updates to the service availability and metric measurements to determine a content delivery server to return.

44. The apparatus of claim 43, wherein each content delivery server in the network has a metric test associated with each service supported by each content delivery server.

45. The apparatus of claim 43, wherein the service metric probe periodically performs metric tests on content delivery servers within the service metric probe area, and wherein the service metric probe records metric results from the periodic tests.

46. The apparatus of claim 43, wherein the latency probe calculates a round trip time for sending a packet to a client to obtain the latency value, and wherein round trip time tests that the latency probe performs, includes any of: PING, UDP Reverse Name lookup, or UDP Packets to high number ports.

47. The apparatus of claim 43, wherein when the latency probe sends a UDP Packet probe to high number ports that fail, the latency probe resends the UDP Packet probe with a low TTL

number and increments the TTL number until failure occurs, a last successful TTL number indicates partial latency data.

48. The apparatus of claim 43, wherein the service metric probe sends an update to all DNS servers in the network that consists of all tests since a last update.

49. (Currently Amended) The apparatus of claim 43, wherein the latency probe updates DNS servers with [[a]] client[[s']] latency data.

50. The apparatus of claim 43, wherein a DNS server uses latency data updates from the latency probe to determine a closest content delivery server to a client.

52. The apparatus of claim 43, wherein the service metric probe sends a packet request to a content delivery server and receives, in response, a packet containing various metrics of the content delivery server, and wherein the service metric probe combines the content delivery server's metrics to arrive at a load metric which is sent to at least one DNS server.

53. (Currently Amended) An apparatus for a determining server performance metrics in a network, comprising:

service metric probe resident on a server for determining service availability and metric measurements of types of services provided by a content delivery server;

latency probe resident on a server for determining a latency value for various servers within the network;

wherein the service metric probe sends an update to all DNS servers in the network that consists of all service availability and metric measurements since a last update; and

wherein the latency probe updates the DNS servers with client[[s']] latency data; and

wherein, for a given DNS name, a DNS server uses updates to the latency values and updates to the service availability and metric measurements to determine a content delivery server to return.

54. The apparatus of claim 53, wherein the service metric probe consults a configuration file containing each DNS name in the service metric probe area and any set(s) of services associated with each DNS name, and wherein the services include any of: HTTP, HTTPS, FTP, streaming media, or generic SNMP.

55. The apparatus of claim 53, wherein the latency probe calculates a latency value from the latency probe location to a client's location.

56. The apparatus of claim 53, wherein each content delivery server in the network has a metric test associated with each service supported by the content delivery server.

57. The apparatus of claim 53, wherein the service metric probe periodically performs metric tests on content delivery servers within the service metric probe area, and wherein the service metric probe records metric results from the periodic tests.

58. The apparatus of claim 53, wherein the latency probe calculates a round trip time for sending a packet to a client to obtain a latency value, and wherein round trip time tests that the latency probe performs, includes any of: PING, UDP Reverse Name lookup, or UDP Packets to high number ports.

59. The apparatus of claim 53, wherein when the latency probe sends a UDP Packet probe to high number ports that fail, the latency probe resends the UDP Packet probe with a low TTL number and increments the TTL number until failure occurs, a last successful TTL number indicates partial latency data.

60. The apparatus of claim 53, wherein a DNS server uses the latency data updates to determine a closest content delivery server to a client.

62. The apparatus of claim 53, wherein the service metric probe sends a packet request to a content delivery server and receives, in response, a packet containing the various metrics of the content delivery server, and wherein the service metric probe combines the content delivery server metrics to arrive at a load metric which is sent to the DNS servers.